JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1999, **71**, 195–214 NUMBER 2 (MARCH)

EQUIVALENCE CLASS ESTABLISHMENT, EXPANSION, AND MODIFICATION IN PRESCHOOL CHILDREN

RICHARD R. SAUNDERS, KATHLEEN M. DRAKE, AND JOSEPH E. SPRADLIN

UNIVERSITY OF KANSAS

Preschool children were taught four two-choice match-to-sample conditional discriminations with 10 arbitrary visual stimuli. For 6 participants, 2 of the 10 stimuli served as the sample, or conditional, stimuli in all discriminations. For 5 additional participants, the same pair of stimuli served as the discriminative, or comparison, stimuli in all discriminations. Equivalence classes were established with more participants in the latter group, replicating prior research with participants with retardation. Four participants, in whom equivalence classes were established and who were available for further participation, were exposed to new conditional discriminations without trial-by-trial feedback and involving some novel and some familiar stimuli. Consistent conditional responding was observed, and tests for inclusion of the novel stimuli in the original classes showed class expansion. Training to reverse the unreinforced conditional performances produced a reversal of class membership in 3 of 4 participants, an outcome not consistent with other studies. The results are discussed with respect to the interaction of class structure and size.

Key words: stimulus equivalence classes, equivalence relations, unreinforced conditional selection, matching to sample, conditional discrimination, preschool children

The study of stimulus equivalence classes typically involves teaching participants a series of conditional discriminations and then testing to determine if new conditional discriminations emerge that were not directly trained (e.g., M. H. Dixon & Spradlin, 1976; Sidman & Tailby, 1982; Spradlin, Cotter, & Baxley, 1973). Some arrangements of test trials reveal whether the relations between the conditional stimuli and discriminative stimuli in the discrimination are conditional relations only (Carter & Werner, 1978) or equivalence relations (Sidman & Tailby, 1982). Equivalence relations have the properties of reflexivity, symmetry, and transitivity. Conditional identity-matching performances indicate the property of reflexivity. Symmetry is demonstrated when the roles of the conditional and discriminative stimuli are reversed and the stimulus–stimulus relations are not affected. Transitivity is demonstrated if two conditional discriminations have stimuli in common and transfer of stimulus control across discriminations is demonstrated as a function of the linking stimuli. Any set of stimuli mutually related by equivalence relations may be described as an equivalence class (Sidman & Tailby, 1982).

Equivalence classes are readily established through conditional discrimination training in persons with retardation and in typically developing children and adults. In a preponderance of these studies, the conditional discriminations have been taught with the match-to-sample (MTS) training paradigm (e.g., Sidman, 1971; Sidman & Tailby, 1982; Spradlin & Saunders, 1986). To test for the relational properties of stimulus equivalence, a minimum of two conditional discriminations must be taught with two stimuli in common to both discriminations. For example, in the first conditional discrimination, in the presence of Sample Stimulus A1 (and not A2) responding to Comparison B1 (and not B2) is correct and in the presence of A2 (and not A1) responding to B2 (and not B1) is correct. In the second conditional discrimination, for example, A1 and A2 are again the sample stimuli and two new stimuli, C1 and C2, become the comparison stimuli. Thus, through A1, Stimuli B1 and C1 could be related and, similarly, B2 and C2 could be related through A2. Although the initial train-

Preparation of this article was supported in part by Federal NICHD Grants 5-P30HD02528 and 1-P01HD18955 to the Schiefelbusch Institute for Life Span Studies, University of Kansas. Computer software for the research was created by Judith Wachter and Richard Saunders. Kathleen Drake is now affiliated with Parsons State Hospital and Training Center, Parsons, Kansas. Julie McEntee and Gina Green provided comments and editorial suggestions that facilitated preparation of the manuscript.

Requests for reprints should be sent to Richard R. Saunders, Parsons Research Center, P.O. Box 738, Parsons, Kansas 67357 (E-mail: rrsaun@parsons.lsi.ukans.edu).

ing could involve additional discriminations, most studies of stimulus equivalence have employed only the minimum two conditional discriminations. The subsequent initial tests for three-member classes usually produce positive results when the standard training and testing protocols have been employed (see Green & Saunders, 1998).

In studies attempting to establish initial classes larger than three stimuli in adults with mild or moderate retardation, however, results have been equivocal (Drake & Saunders, 1987 as cited in K. Saunders, Saunders, Williams, & Spradlin, 1993; R. Saunders, Wachter, & Spradlin, 1988; Spradlin & Saunders, 1986). Across these studies, when the training employed the same pair of sample (conditional) stimuli for four pairs of discriminative (comparison) stimuli, equivalence classes were established in only 1 of 7 participants. In the one successful case, each class contained one sample stimulus and four comparison stimuli. In contrast, when the training employed one set of comparison stimuli across conditional discriminations with four pairs of sample stimuli, equivalence classes were established in 6 of 6 participants. These classes contained one comparison stimulus and four sample stimuli per class. Thus, the degree to which equivalence classes were demonstrated may have depended on the *training structure,* that is, on how stimuli in the potential classes were linked (Fields & Verhave, 1987; Green & Saunders, 1998). The two structures just described have been referred to as respectively as one to many (Urcuioli & Zentall, 1993) and many to one (Urcuioli, Zentall, Jackson-Smith, & Steirn, 1989) or sample as node (SaN) and comparison as node (CaN) (K. Saunders et al., 1993). A node refers to a stimulus in a potential class that provides the relational linkage among two or more other stimuli (Fields, Verhave, & Fath, 1984).

Training-structure differences also have been reported with participants without mental retardation. Barnes, using a single-test protocol, found that equivalence relations were established in 9 of 10 normal adult participants taught with a CaN structure, but these relations were established in only 4 of 10 participants taught with a SaN structure (Barnes, 1992, reported in Barnes, 1994). Unfortunately, class size was not reported by Barnes (1994). Recently, Fields, Hobbie, Adams, and Reeve (in press) also reported training-structure effects, but of a somewhat different nature. Fields et al. compared training with CaN and SaN structures in normal adults across two potential class sizes: five and seven stimuli per class. Equivalence classes were established eventually in all participants, but with the seven-member classes, the relations emerged more quickly following CaN training than following SaN training. K. Saunders et al. (1993) reported somewhat similar results with typically developing children 8 to 14 years old. Following SaN training leading to classes of five or more members, equivalence relations did emerge, but in 2 of the 3 participants they did so only gradually.

In contrast to research conducted with older participants, research conducted with young children has not explicitly examined the effects of training structure or class size. Instead, initial testing follows training with only the minimum number of prerequisite conditional discriminations (Barnes, Browne, Smeets, & Roche, 1995; Barnes, McCullagh, & Keenan, 1990; Devany, Hayes, & Nelson, 1986; Lazar, Davis-Lang, & Sanchez, 1984; Pilgrim, Chambers, & Galizio, 1995; Sidman, Kirk, & Willson-Morris, 1985; Sidman, Willson-Morris, & Kirk, 1986). Thus, testing for two three-member classes has been the standard with this population. Equivalence relations have been demonstrated in nearly all typically developing children and some with mental retardation between the ages of 2 and 7 years. None of these studies employed CaN procedures; all employed SaN procedures except that by Lazar et al. (1984), who employed what has been referred to as a linear series structure (e.g., Green & Saunders, 1998; K. Saunders et al., 1993).

Another age-related difference reported in the literature concerns the effects of reversal training. Michael and Bernstein (1991) reported that following the establishment of two three-member equivalence classes in 4- and 5-year-old participants, training to reverse two previously trained relations in one of the conditional discriminations disrupted equivalence test performances. That is, the test performances were not consistent with the previous equivalence classes, nor were they consistent with the reversal training. Performance on the unreversed relations in the second baseline conditional discrimination was also disrupted by the reversal training. Pilgrim et al. (1995) found that reversal training on a discrimination following establishment of equivalence classes in 5- to 7-year-old children disrupted responding on tests for equivalence. Test performances were not consistent with either the equivalence classes based on the former contingencies or on the reversed contingencies.

Both the Michael and Bernstein (1991) and the Pilgrim et al. (1995) reversals were conducted on three-member equivalence classes; that is, classes based on two conditional discriminations. In contrast, Spradlin, Saunders, and Saunders (1992) reported that for 2 typically developing children (8 and 12 years old) exposed to training leading to two five-member classes, reversal of one conditional discrimination did not alter equivalence class organization. Spradlin et al. hypothesized that the larger the equivalence classes established by training, the less susceptible they are to disruption of a single relation. With college students, Pilgrim and Galizio (1995) found that after training two four-member, and later, five-member classes, reversal of particular conditional discriminations in the training structure did not alter class organization, despite the fact that performances on symmetry tests were consistent with the reversal training.

R. Saunders, Saunders, Kirby, and Spradlin (1988, Experiment 2) reported results from a somewhat different preparation, but with related results with adolescents and adults with mental retardation. In this experiment, participants were exposed to two novel trial types without trial-by-trial feedback. In the novel trials, a pair of comparison stimuli (nodal stimuli) from previously established four-member equivalence classes were presented in the presence of either of two samples from another pair of equivalence classes. These participants responded consistently to each comparison in the presence of a different sample, and it was later demonstrated on tests that the classes from which these samples and comparisons were drawn had merged. Next, participants were exposed to reversal training on the relations formed without trial-by-trial feedback. That is, if an unreinforced response had been to Comparison A1 (in one class) in the presence of Sample F1 (in another class) and to A2 in the presence of F2, responding to A2 in the presence of F1 was reinforced, and vice versa. When this reversal training established stable responding (which was achieved only after numerous training sessions), tests for equivalence were readministered. Maintenance of the original merged classes was observed for 2 of the 3 participants, despite the reversed responding on the class-linking relations. For the 3rd participant, maintenance of the original class merger was seen, but only on a particular subset of test trials. That subset included only those test trials that presented the F stimuli, either as the comparisons or as a sample. For this participant, the reversal training reversed the class membership of the F stimuli in the larger classes but did not alter the remaining organization of the larger classes.

Thus, to date, stimulus equivalence research involving young children as participants has differed from that of research involving older participants. First, training structure differences have been reported with adults but not children. Second, reversal training has disrupted class-consistent responding to a greater degree in children than in adults. It is possible that both of these age-related differences are related to differences in the size of the classes established in training. Would training leading to larger initial classes reveal training-structure differences in young children? Would exposing young children to reversal training following the development of larger classes show the equivalence class stability shown previously by older participants with larger classes?

To address these questions, Experiment 1 was designed to compare CaN and SaN procedures with larger initial prospective classes (five members) in typically developing children. Participants were selected whose chronological ages generally corresponded to the mental ages of participants with retardation who have shown training-structure differences (Drake & Saunders, 1987, as cited in K. Saunders et al., 1993; R. Saunders, Wachter, & Spradlin, 1988; Spradlin & Saunders, 1986). The participants also were as young or younger than previous typically developing participants in equivalence research. Experiment 2 was designed to determine whether larger initial equivalence classes would be re-

Table 1

Participant characteristics: age, training procedure assignment, and the experiments in which they participated listed from youngest to oldest in each group.

Participant	Age	Training procedure	Experiment participation
Tina	3 years 5 months	SaN	1
Mary	3 years 10 months	SaN	1
Rita	4 years 1 month	SaN	1
Rob	4 years 6 months	SaN	1
Ann	5 years 3 months	SaN	1, 2
Beth	5 years 3 months	SaN	1
Chad	3 years 7 months	CaN	1
Mike	3 years 11 months	CaN	1, 2
Alice	4 years 3 months	CaN	1
Chet	5 years 3 months	CaN	1, 2
Jenny	5 years 4 months	CaN	1, 2

sistant to disruption by reversal training with young children.

EXPERIMENT 1

METHOD

Participants

Eleven preschoolers attending the Edna A. Hill Child Development Center at the University of Kansas participated. The mean age of participants was 4 years 5 months at the onset of Experiment 1. Table 1 lists the participants' ages and the experiments in which they participated.

Apparatus

Each participant was seated in a chair in front of a small table that supported an Apple IIE® computer, floppy diskette drives, a monochrome (green phosphorous) monitor, and a printer. The face of the monitor was covered with a Personal Touch Touchscreen® that served as an interface between the computer and the participant's responses. The touchscreen acknowledged touches when the pressure applied caused the film-like surface to make contact with the rigid plastic backing. During the experimental sessions, the monitor displayed stimuli on the screen in three zones, each measuring approximately 5 cm by 5 cm. Touching the screen within these zones advanced the trial sequence and resulted in programmed consequences. Stimuli were displayed on the monitor screen using the computer's high-resolution graphics. On the table to the left of the participant was a small cup into which the experimenter could place plastic tokens according to the schedule of reinforcement in effect.

Procedure

The general procedure for each session was similar to that described in R. Saunders, Wachter, and Spradlin (1988) and R. Saunders, Saunders, Kirby, and Spradlin (1988). Each trial began with a sample stimulus displayed in the center zone. A touch within that zone resulted in the concurrent display of two comparison stimuli in the zones to the right and left of center. Further touches in the center zone had no programmed consequences, although the sample stimulus remained displayed. A touch in the zone for either comparison stimulus resulted in programmed consequences (see below), the removal of all displays during a 2-s intertrial interval (ITI), and the presentation of the next sample stimulus.

During training, selection of the comparison stimulus designated as correct resulted in a 1-s computer-generated auditory jingle that imitated the musical phrase "ta da, ta daa" that often prompts fans at sporting events to yell, "charge." The jingle was accompanied by the manual delivery of a token into the cup by the experimenter. Incorrect responses resulted in a buzzer sound of approximately 1-s duration. On test trials, to be described later, and on training trials with no programmed consequences, selections of comparison stimuli resulted in the removal of all stimulus displays and advancement to the next trial following the 2-s ITI.

At the end of each session, the tokens were traded for a small toy, sticker, or hair ribbon, for example, from an array of such items. The participant could choose any toy from the array, regardless of the number of tokens received during the session.

Overview of Training and Testing

One at a time, each participant accompanied the experimenter from the preschool to the laboratory setting each weekday and participated for up to 15 min each day, permitting one or two experimental sessions during that time. Experiment 1 had three phases.

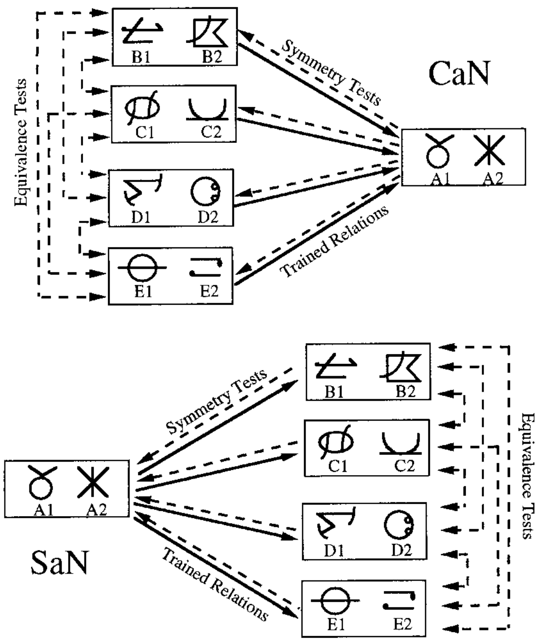


Fig. 1. Schematics of the training structures employed in Experiment 1. The upper panel depicts the CaN structure leading potentially to two classes of five members each. The lower panel depicts the SaN structure leading potentially to two classes of five members each. Solid lines indicate trained relations and point from sample stimuli to comparison stimuli; dashed lines indicate tests for emergent relations. The stimuli are shown with alphanumeric labels for organizational purposes. The numeral in each label indicates membership in a stimulus equivalence class (Class 1 or Class 2) that should be the result of the training procedures employed. The letters in each label indicate how the stimuli were paired for introduction to the participants.

The first phase produced identity-matching performances. Although identity matching conducted prior to arbitrary MTS training is not considered an appropriate test for the property of reflexivity (R. Saunders & Green, 1992; Sidman, 1994), this phase may be considered to be pretraining in which the participant gained experience with conditional discrimination problems and with the apparatus. This phase also provided an initial test of whether the participant would develop conditional discrimination performances in an MTS task. Four arbitrary conditional discriminations were taught in Phase 2, under one of two training procedures shown in Figure 1 and described below. Phase 2 was completed when the experimenter-designated correct performances were maintained under a reduced probability of reinforcement (.75). In Phase 3, two types of tests were administered, as shown in Figure 1. In a test for the property of symmetry, stimuli that were sample stimuli in training were presented as comparisons in the presence of the former comparisons as samples (i.e., sample-comparison role reversal). Following CaN training, in tests for the combined properties of symmetry and transitivity or direct tests for equivalence (Sidman, 1990; Sidman, Wynne, Maguire, & Barnes, 1989), former sample stimuli were presented as comparison stimuli in the presence of other sample stimuli. For direct tests of equivalence following SaN training, former comparison stimuli were presented as sample stimuli for other former comparison stimuli.

Phase 1: Identity Matching

Each participant received 16 trials of identity matching per session, with stimuli similar to those shown in the schematic in Figure 1, until criterion was met. Each session was constructed such that a stimulus that served as the nonmatching, or incorrect, comparison on one or more trials also served as a sample and matching, or correct, comparison on other trials; thus, the identity-matching training was arranged to produce conditional identity matching (Green & Saunders, 1998; Mackay, 1991). Criterion for completion of Phase 1 was 90% or better correct for two consecutive sessions or 100% correct for one session. The participant was instructed on the first trial of the first session to "Touch the one that goes with the one in the middle." The stimuli used in this phase were not used again during this experiment. Table 2 shows that performances met criterion in one to eight sessions.

Phase 2: Arbitrary Conditional Discrimination Baseline Training

Arbitrary conditional discrimination trials were conducted using the stimuli shown in Figure 1. Phase 2 training began with the BA or AB conditional discrimination (i.e., B1A1 and B2A2, or A1B1 and A2B2) depending on the training group to which the participant was assigned. Six participants were assigned to the SaN group, in which the A stimuli served as the sample stimuli in all four discriminations and the B, C, D, and E stimuli served as pairs of comparison stimuli, as

Table 2

Sessions to criterion for each participant in Phases 1 and 2, with participants listed from youngest to oldest in each group.

Participant	Training procedure	Phase 1	Phase 2
Tina	SaN	8	11
Mary	SaN	2	74
Rita	SaN	2	12
Rob	SaN	1	19
Ann	SaN	1	14
Beth	SaN	1	13
Chad	CaN	2	36
Mike	CaN	1	48
Alice	CaN	1	37
Chet	CaN	2	22
Jenny	CaN	1	45

shown in Figure 1. Five participants were assigned to the CaN group, in which the A stimuli served as the comparison stimuli throughout training, also shown in Figure 1. Participants were assigned to the two groups systematically by age to produce roughly comparable age ranges within each group. The participants taught with SaN procedures were instructed on the first trial, ''When the spider comes up, you touch the flag,'' and on the second trial, ''When the apple comes up, you touch the elbow.'' The same instructions were repeated on the next two trials. The participants taught with CaN procedures were given the same instructions, but with the stimulus names reversed in each sentence. The four trials containing instructions replicated the procedures used in the research on persons with mental retardation (e.g., R. Saunders, Wachter, & Spradlin, 1988). No further instructions regarding stimulus names were given throughout the remainder of the experiment.

When performance met criterion on the BA or AB discrimination training, that training was discontinued and training on the CA or AC discriminations began. Separate sessions of training on the DA or AD and EA or AE discriminations were conducted similarly. When a participant's performance on the EA or AE discrimination met criterion, training on a mix of the BA, CA, DA, and EA or AB, AC, AD, and AE discriminations began. These sessions contained an approximately equal number of training trials of each discrimination. When a participant's performance on the mix of the four discriminations met criterion, sessions of the same mix were conducted, but with four trials without scheduled consequences (no sound or tokens delivered). Participants were told that on some trials, ''The computer will be silent, but continue working.'' Chet, Beth, Ann, and Tina completed the training in the sequence described above. The remaining participants required one or more sessions of refresher training on individual discriminations to establish stable performances with the mixed discriminations and reduced feedback.

Phase 3: Probes for the Properties of Equivalence and Symmetry

Following training, a series of sessions was conducted to test for equivalence (combined tests of symmetry and transitivity) and separately for the property of symmetry. Figure 1 shows the stimulus–stimulus relations tested. Each test session contained four probe trials with a selection of two probes from prospective Class 1 and two from prospective Class 2. On all test trials, the comparison stimuli were presented in the same stimulus pairs that had been presented in training (e.g., if C1 was the correct comparison stimulus, C2 was the second comparison). Responses on test trials had no programmed consequences. The test trials were interspersed among the original training trials unsystematically, except that test trials were always separated by one or more training trials. The participant was instructed as in Phase 2 regarding the computer's silence. At the end of each session, regardless of test performance, the experimenter added four tokens to the cup as a quantity; that is, the tokens were placed in the cup without counting them out to the participant. If a participant's performance on the baseline trials fell below 90% correct in any test session, the testing sequence was interrupted and Phase 2 training sessions were resumed until the criterion for initiating the test sequence was met again.

Four sessions with four equivalence test trials were conducted first, followed by two sessions with four test trials for the property of symmetry, and then two additional sessions with four equivalence test trials. Symmetry testing was redundant because symmetry is evaluated in each equivalence test trial (simultaneous tests for transitivity and symmetry). Tests for symmetry only were included, however, to determine whether the property of symmetry might be shown when transitivity was not. Across the eight test sessions, each possible test trial type was presented only once, although in tests for equivalence each bidirectional stimulus–stimulus relation was tested twice (e.g., B1C1 and C1B1). Because the test series was lengthy, multiple tests of each trial type were not planned if the initial test series indicated the establishment of stimulus equivalence classes.

To determine appropriate criteria for interpreting whether one-trial unreinforced test results showed the establishment of stimulus equivalence classes, we extrapolated from Sidman's discussions of potential problems with two-choice matching (Carrigan & Sidman, 1992; Johnson & Sidman, 1993; Sidman, 1987). For example, Sidman (1987) reasoned that, in two-choice matching, relatively high scores such as 75% class-consistent responding could be misleading. With repeated testing of pairs of stimulus–stimulus relations (e.g., A1B1, A2B2), 75% class-consistent responding could be a function of (a) above-chance stimulus control by both samples, (b) stimulus control by one sample (e.g., A1) and chance responding in the presence of the second sample, or (c) stimulus control by one sample (e.g., A1) and position responding in the presence of the second sample. Thus, Sidman urged adoption of higher overall stringency in establishing criteria for data interpretation.

Although training multiple discriminations leading to large classes reduces the possibility of incorrect inferences from scores of 75%, for example (Green & Saunders, 1998), the present criteria were nevertheless set to be responsive by analogy to Sidman's (1987) reasoning. In the present study, 83% (20 of 24) or higher equivalence-indicative performances and 87% (seven of eight) or higher symmetry-indicative performances were accepted as demonstration of the relational properties tested. That is, participants who responded in a manner consistent with the experimenter-defined classes on more than 83% of equivalence test trials and 87% of symmetry test trials were not tested again. We designed the criteria to compare control by all experimenter-defined Class 1 samples (on tests) with control by all experimenter-defined Class 2 samples. For both types of tests, the criterion was made sufficiently stringent so that it could not be met when samples from one class controlled responding and samples from the second class did not. Individual test results also were reviewed for specific error patterns and are discussed below.

One participant had an equivalence test performance above 83% but had a symmetry test performance below 87%, and 1 participant had a symmetry test performance above 87% but an equivalence test performance below 83%. These participants were retested under the same test conditions as the first test series. Three other participants had performances on tests that met neither criterion. Retesting without trial-by-trial feedback for all trials was chosen as the retest method for these 3 participants because a similar procedure had induced class-consistent performances in 1 participant in R. Saunders, Wachter, and Spradlin's (1988) study, who initially had test performances that did not differ from chance responding. These participants were returned to Phase 2 training conditions, and the probability of scheduled consequences was reduced to zero. The participants were instructed that the computer would be "silent all the time," but that the experimenter would keep track of their performances and deliver tokens at the end of the session. At the end of each session, the participant received one token for each correct response on training trials and a token for each test trial, regardless of the participant's response on test trials. The tokens were delivered as a quantity in a cup, however, so the participant was not made aware of their relationship to types of trials from which they derived nor the total earned. When performance on the training trials without trial-by-trial feedback met criterion for testing a second time, the test series was repeated.

RESULTS

As shown in Table 2, the number of sessions required by SaN participants to complete Phase 2 training varied from 11 to 74, with a mean of 23.8 and a median of 13.5. Figure 2 shows the results of the tests and retests in Phase 3 for the participants in the SaN group. For only Rita and Ann were two five-member equivalence classes established

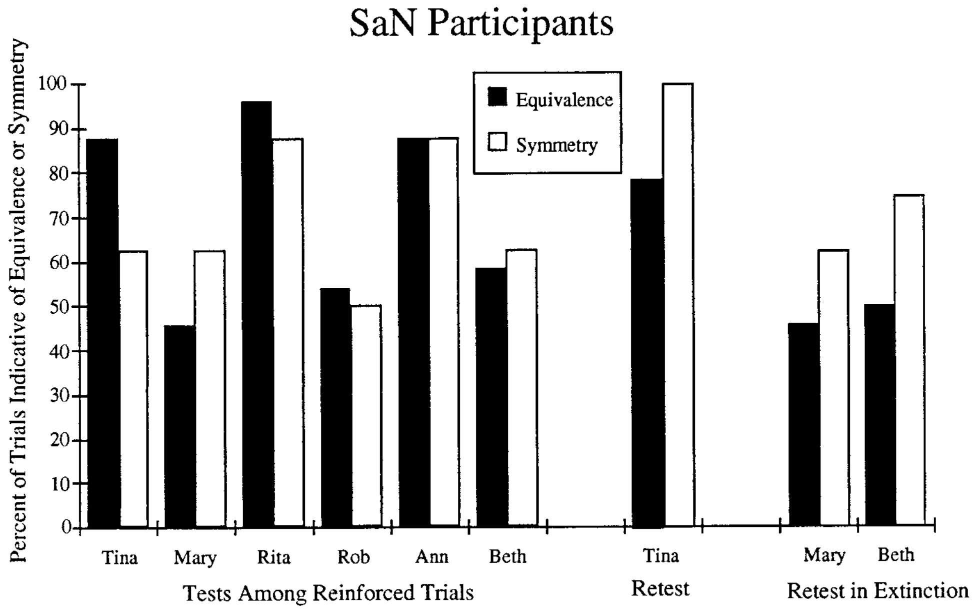


Fig. 2. Percentage of test trials indicative of equivalence and symmetry for the 6 participants in Experiment 1 who received training in the SaN structure. From left to right, the first six sets of bars show the results of the first test series with each participant, with the participants ordered from youngest to oldest. The sets of bars to the right show the results of the second test series with Tina, Mary, and Beth.

in the first set of tests. Rita and Ann had one and three class-inconsistent responses, respectively. Each of Ann's involved a different bidirectional stimulus–stimulus relation, and these were spread across both classes (C1B1, D1C1, D2E2). Responding on equivalence and symmetry test trials by Mary, Rob, and Beth was unsystematic. Despite retests without trial-by-trial feedback, neither Mary's nor Beth's performances met the criteria; Rob was unavailable for further testing. Tina's performance met the criterion for equivalence but not symmetry. Her retest performance (shown in Figure 2) shows the same outcome, but with a reversal of which test type on which the criterion was met. In the first equivalence test series, Tina made three class-inconsistent responses with stimulus–stimulus relations D2E2, B2C2, and C2D2—all different bidirectional relations, but all from Class 2. During retesting, her class-inconsistent responses were with stimulus–stimulus relations E2B2, E2D2, B2C2, C2D2, and B2D2. Thus, in the retest, Tina showed class-consistent responding in the presence of Class 2 samples on 7 of 12 trials and on 12 of 12 trials in the presence of Class 1 samples.

As shown in Table 2, the number of sessions required by CaN participants to complete Phase 2 training varied from 22 to 48, with a mean of 37.2 and median of 37. As shown in Figure 3, equivalence relations were established in 3 of the 5 participants trained with the CaN procedures. Chad, Mike, and Alice each had one class-inconsistent response in the equivalence tests. Chet, whose performance met criteria only on the symmetry tests, was retested with trial-by-trial reinforcement on baseline trials and showed 100% equivalence-indicative responding on the equivalence tests. Due to experimenter error, the symmetry tests were not repeated; symmetry was evaluated, however, in each equivalence test trial. Jenny's performance did not meet the criteria on either test type in the first test series. A retest without reinforcement on any trial, however, produced performances that met the criteria for equivalence. Thus, equivalence relations were eventually established in all CaN trained participants. Chet had no class-inconsistent responses during retesting; Jenny had one such response in retesting.

Discussion

Tina's response pattern in the equivalence retest is indicative of pattern (b) described by

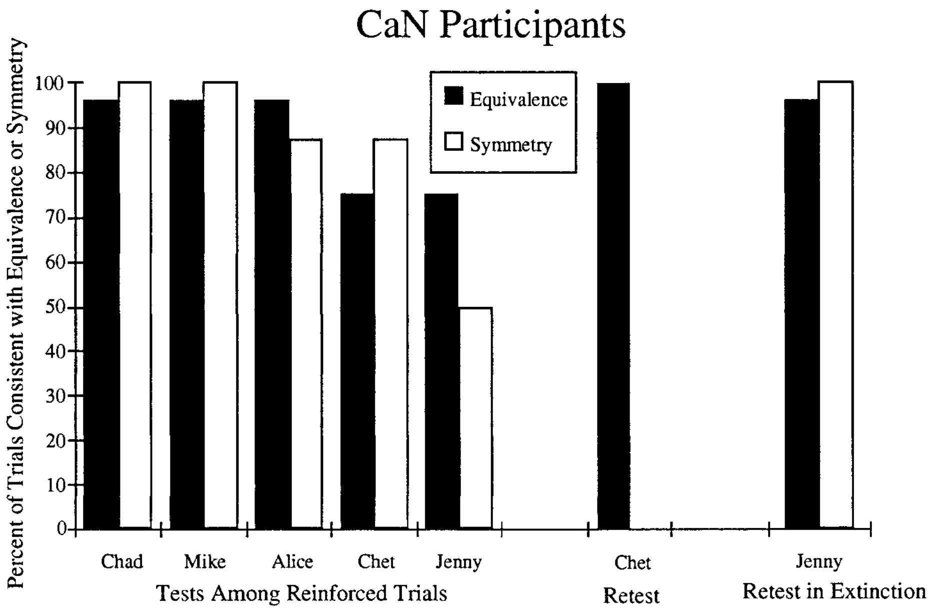


Fig. 3. Percentage of test trials indicative of equivalence and symmetry for the 5 participants in Experiment 1 who received training in the CaN structure. From left to right, the first five sets of bars show the results of the first test series with each participant, with the participants ordered from youngest to oldest. The sets of bars to the right show the results of the second test series with Chet and Jenny.

Sidman (1987). Her responses in the presence of samples from Class 2 were unsystematic, sometimes responding to a Class 1 comparison and about equally often responding to a Class 2 comparison. Although Class 1 samples consistently produced responses to Class 1 comparisons and her overall percentage of class-consistent responses was high, her overall performance met neither our nor Sidman's criterion for equivalence. Thus, a conclusion that equivalence classes were established in Tina is not justified, despite the 100% class-consistent performance in the symmetry retest. Therefore, equivalence classes were established in only 2 of 6 participants trained with SaN procedures. In contrast, equivalence classes were established eventually in all CaN trained participants. These differences in test performances as a function of training history in our preschool participants are similar to the differences found in previous studies with adolescents with mild or moderate retardation. This similarity suggests that young children and persons with mild or moderate retardation may have learning and performance characteristics in common, at least with respect to the establishment of larger (five-member) equivalence classes. The results with the SaN trained participants, however, contrast with the results of previous research with young children with three-member classes. A recent conceptual and methodological analysis of stimulus equivalence offers one explanation for the reported differences.

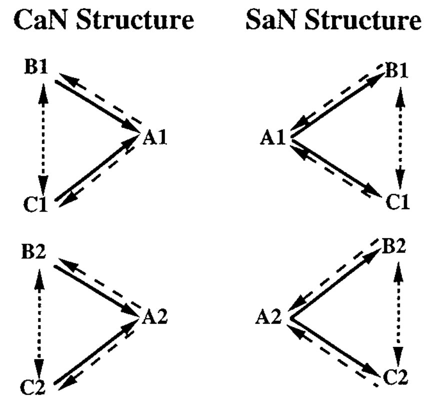
Extending an analysis by K. Saunders et al. (1993) and Spradlin and Saunders (1986), R. Saunders (1997)[1] mapped how training with different structures may require differing numbers of simple discriminations among the stimuli in the experiment. Specifically, R. Saunders explained that training in the CaN structure requires a simple discrimination (simultaneous or successive) of each stimulus from every other stimulus in the training if the typical criterion for testing is to be met

[1] Saunders, R. R. (1997). *Stimulus equivalence as a contingency element.* Paper presented at the European meeting of the Experimental Analysis of Behaviour Group, Dublin, Ireland. Substantial contributions to the analytical content of this paper were made by Gina Green and were acknowledged in the presentation. Her contributions were based in part on her earlier paper: Green, G. (1997, May). *What's in a discrimination.* Paper presented at the annual meeting of the Association for Behavior Analysis, Chicago.

(e.g., +90% correct). With the SaN structure, only some of the possible simple discriminations are required for the trained performances to meet an equivalent criterion. R. Saunders' analysis, hereafter referred to as the discrimination analysis, assumes that (a) simultaneous MTS procedures are used, (b) baseline conditional discrimination problems are presented together within sessions at some point before testing begins, and (c) during training and testing, the commonly employed paired-comparison method of presenting comparison stimuli is used (as explained below). These assumptions were met by the protocol of the current experiment.

The discrimination analysis is relevant to the training-structure results here and those found previously with participants with mental retardation in two ways. First, performances on tests for the properties of equivalence are dependent on each stimulus in the tests being discriminated from every other stimulus in the tests. Second, the number of simple discriminations not required during SaN training, but required on subsequent tests, increases as the intended class size increases. Figure 4 shows schematics of the three-member classes arising from CaN and SaN training with the minimum number of conditional discriminations. This figure also shows that the trial types for tests for equivalence following training with either of the two structures are the same. In CaN training, the stimuli that will be involved in the test trial types for equivalence must be discriminated successively from one another if the participant is to perform in accordance with the training contingencies. That is, either B1, C1, B2, or C2 is always the sample on every training trial and must be discriminated successively from the other samples. This requirement for discriminations among the samples insures that B1 is discriminated from B2 and C2 and vice versa and that C1 is discriminated from B2 and C2 and vice versa. These discriminations are essential to performing correctly in the conditional discriminations that constitute the tests for equivalence, as shown in Figure 4.

In contrast, during SaN training, B1 must be discriminated as a comparison stimulus from B2 and, similarly, C1 must be discriminated from C2. Prior to testing, however, SaN training with the paired-comparison procedure does not require B1 to be discriminated from C1 or C2 because in training B1 is always presented with B2 and never with C1 or C2 in the same trial (Barnes, 1994; Spradlin & Saunders, 1986). Similarly, B2 does not have to be discriminated from C1 or C2. Thus, training leading to two three-member classes with a SaN structure and paired-comparison training does not require four of the simple discriminations needed for consistent performances on the tests for equivalence. Of these four discriminations, two are particularly critical: B1 from C2 and B2 from C1. That is, a failure to discriminate between stimuli from within a potential class (e.g., B1 from C1) is not as likely to lead to negative test results as a discrimination failure between stimuli from different potential classes.



Equivalence Trial Types for Both Structures

Samples	Comparisons
B1	*C1 C2
B2	C1 *C2
C1	*B1 B2
C2	B1 *B2

Fig. 4. Schematics of the CaN and SaN training structures, each leading potentially to two classes of three stimuli each. Solid arrows indicate trained relations, dashed-line arrows indicate tests for the relational property of symmetry, and the double-pointed dotted-line arrows indicate tests for the combined relational properties of symmetry and transitivity or direct tests for equivalence. The trial types for direct tests for equivalence are the same for both structures and are shown below the schematic. Asterisks indicate the experimenter-designated correct comparison (for equivalence-consistent responding).

Despite the lack of a requirement to learn the discriminations during training, some discriminations may be learned anyway. Further, when the number or proportion of unrequired discriminations is small, mixing test trials among baseline trials, as is common, may result in rapid acquisition of the remaining discriminations during testing. This outcome seems less probable, however, when the number or proportion of unrequired discriminations is larger. For example, SaN training leading to two five-member classes, as in the present experiment, requires only 21 of the 45 simple discriminations that are needed on the subsequent tests for equivalence, and 12 of the 24 remaining discriminations involve discriminations between stimuli from different experimenter-designated classes. A reasonable hypothesis is that differences in requirements for simple discrimination acquisition imposed by the larger classes produced the differences in test results with persons with retardation in prior research and with the present participants.

One reason why SaN training may have resulted in more frequent equivalence than expected on the basis of the simple discriminations required in training stems from the instructions used in the experiment. To precisely replicate the procedures of R. Saunders, Wachter, and Spradlin (1988), the present participants received four trials with instructions that included stimulus names (e.g., flag, apple) in the first training session of Phase 2. If the experimenter-offered names prompted stimulus naming thereafter, unique naming of each stimulus would result in a successive simple discrimination of each stimulus from every other stimulus. Thus, if naming were employed by any participant trained with either the CaN or the SaN structure, all the discriminations necessary for equivalence-indicative test performances would be acquired prior to testing, despite differences in what the training structures required. Particularly with verbal participants, naming may arise in experiments in which the procedures may prompt (but not necessarily require) naming (McIlvane & Dube, 1996; and cf. K. Saunders, 1989; Sidman & Tailby, 1982). Researchers have also used naming to facilitate conditional discrimination performances (e.g., Dugdale & Lowe, 1990; Eikeseth & Smith, 1992; Lowe & Beasty, 1987). Thus, despite the fact that testing occurred hundreds of trials after the instructions had been given, the instructions could have influenced discrimination learning and, therefore, equivalence and symmetry test performances.

Some support for this interpretation derives from K. Saunders et al.'s (1993) report that across two experiments with adolescent and adult participants with mental retardation, CaN training was more likely to produce equivalence classes if particular instructions were used. The instructions were the same as those employed in the present experiment and the preceding research with participants with mental retardation (Drake & Saunders, 1987, as cited in K. Saunders et al., 1993; R. Saunders, Wachter, & Spradlin, 1988; Spradlin & Saunders, 1986). K. Saunders et al. reported that two five-member equivalence classes were established in 5 of 6 subjects trained with instructions, but these were established in only 4 of 8 participants trained without these instructions. In the experiment reported by R. Saunders, Saunders, Kirby, and Spradlin (1988), two four-member equivalence classes were established in 3 adults with mental retardation following training with CaN procedures that also included these instructions.

Arntzen and Holth (1997) recently reported results in conflict with the preceding findings with respect to training-structure differences. Three three-member equivalence classes were established in 10 of 10 normal adult participants taught with a SaN structure but were established in only 7 of 10 participants taught with a CaN training structure. One consideration in interpreting Arntzen and Holth's results is that testing was done in blocks of test trials; that is, tests trials were not intermixed with training trials. Arntzen and Holth also reported that the CaN training structure led to more errors than did the SaN training structure during the final training phase in which both conditional discriminations were mixed in single sessions prior to testing. It is possible that the test differences were a function of differences in the retention of the trained discriminations during testing. Because simple simultaneous discriminations have been found to be easier to teach than simple successive discriminations (Brady & Saunders, 1991; Carter & Eckerman, 1975; Urcuioli et al., 1989), SaN training might have produced better acquisition of the initial conditional discriminations than did CaN training. If the discriminations in SaN training were acquired sooner, then by the time criterion for testing was met, overtraining may have occurred. Spradlin et al. (1992) suggested that the extent of overtraining could be related to the degree of stability in equivalence classes. Thus, despite training to the same criterion prior to testing, Arntzen and Holth's SaN participants may have been better prepared than their CaN participants to retain the baseline discriminations during blocks of test trials.

The Phase 2 training data in the present study appear to bear out this possibility. The mean number of sessions to criterion in Phase 2 following SaN training was only 23.8 (median, 13.5); the mean following CaN training was 37.2 (median, 37). This difference suggests that the SaN conditional discriminations were easier to acquire than the CaN conditional discriminations. When the sessions-to-criterion means in the present study are converted to trials to criterion (for meaningful comparison with other studies), the mean difference is 381 (SaN) to 595 (CaN). Converted to trials to criterion, R. Saunders, Wachter, and Spradlin (1988) reported mean differences of 1,024 (SaN) to 1,449 (CaN). Fields et al. (in press) also reported that acquisition of the CaN discriminations required more training than acquisition of the SaN discriminations. Thus, for preschool children, normal adults, and adolescents with retardation in similar experiments, SaN discriminations have been learned more readily, but have led less quickly or less frequently to equivalence-indicative performances than have CaN discriminations.

EXPERIMENT 2

Some subjects with mental retardation or developmental delay, after being taught a series of conditional discriminations, have performed with few or no errors on subsequently presented new conditional discriminations, even when the new relations could not be derived from previously trained relations (e.g., L. S. Dixon, 1977; K. Saunders & Spradlin, 1990; R. Saunders, Saunders, Kirby, & Spradlin, 1988; Stromer & Osborne, 1982; Wetherby, Karlan, & Spradlin, 1983). R. Saunders, Saunders, Kirby, and Spradlin referred to consistent sample-comparison responding in the absence of trial-by-trial feedback as unreinforced conditional selection. Subjects in whom larger equivalence classes were established in Experiment 1 would have learned four conditional discriminations, thus providing an appropriate history for testing for unreinforced conditional selection. Thus, Experiment 2 was designed to evaluate whether young children would make unreinforced conditional selections. The procedures were arranged so that class expansion could occur based on the unreinforced conditional selections, and equivalence tests were arranged for confirmation. Reversal training, the primary independent variable in Experiment 2, was scheduled next, either on the expanded classes (if expansion occurred) or on the large classes formed in Experiment 1 (if expansion did not occur). Equivalence tests followed reversal training to determine the degree to which such training disrupted or reorganized the classes.

METHOD

Participants and Apparatus

Experiment 2 was conducted with 4 participants from Experiment 1 in whom equivalence classes had been established and who were available for further participation: Ann (a SaN-trained participant) and Mike, Jenny, and Chet (CaN-trained participants). The apparatus was the same as used in Experiment 1.

Phase 1: Maintenance of Trained and Tested Relations Without Trial-by-Trial Feedback

To produce stable performances on test and training trials under conditions of no trial-by-trial feedback and to create a new baseline for testing in subsequent phases, the participants were exposed to (a) retesting for equivalence (but not symmetry) in six sessions in which tokens were delivered following correct responses on training trials only (as in Phase 3, Experiment 1), (b) retesting for equivalence in two sessions without any scheduled consequences, and (c) retesting for equivalence in four sessions without scheduled consequences with eight test trials (increased from four) and eight training trials. The sessions with eight test trials will be referred to in the Results as the final tests of

Phase 1. The participants were instructed that the computer would be silent in (b) and (c) above, but that the experimenter would keep track of their performances and deliver tokens at the end of the session, as described previously. The criterion for maintenance of the training trials was changed to 87.5% correct (seven of eight), and the criterion for maintenance of the equivalence classes was set at 87.5% class-consistent responding (seven of eight).

Phase 2: Unreinforced Conditional Selection with Two Novel Stimuli

In this phase, the participants were exposed to 16-trial sessions with no scheduled consequences in which only two types of trials were presented, FA or AF. Figure 5 shows the training schematics enlarged to add the FA and AF discriminations, respectively. On half of the trials for the CaN-trained participants, the novel stimulus, F1, served as the sample stimulus and the previous comparison stimuli, A1 and A2, served as the comparisons. In the other half of the trials, F2 served as the sample with the A comparisons. For Ann (the SaN-trained participant), the sessions consisted of trials with the original A samples and new F comparisons. The order of trials was unsystematic except that neither type of trial could occur more than three times in succession. Sessions of this type were conducted until responses were consistent for two consecutive sessions. That is, if A1 was usually selected in the presence of F1 and A2 was selected in the presence of F2, criterion was met when this response pattern was maintained on at least 15 of 16 trials in two consecutive sessions. The participants were given 16 tokens at the end of each session, however, regardless of performance.

Phase 3: Tests for Equivalence Between the F Stimuli and the Trained Stimuli

Test sessions consisted of 16 trials without scheduled consequences that included six former training trials, two FA trials (or AF for Ann), four test trials for equivalence among *trained* samples (or comparisons for Ann), and four test trials that tested for equivalence (combined transitivity and symmetry) between the F stimuli and the trained samples (or trained comparisons for Ann). These latter four trials tested for the incorporation of the F stimuli into the existing equivalence classes (see Figure 5). Four test sessions were conducted in this phase.

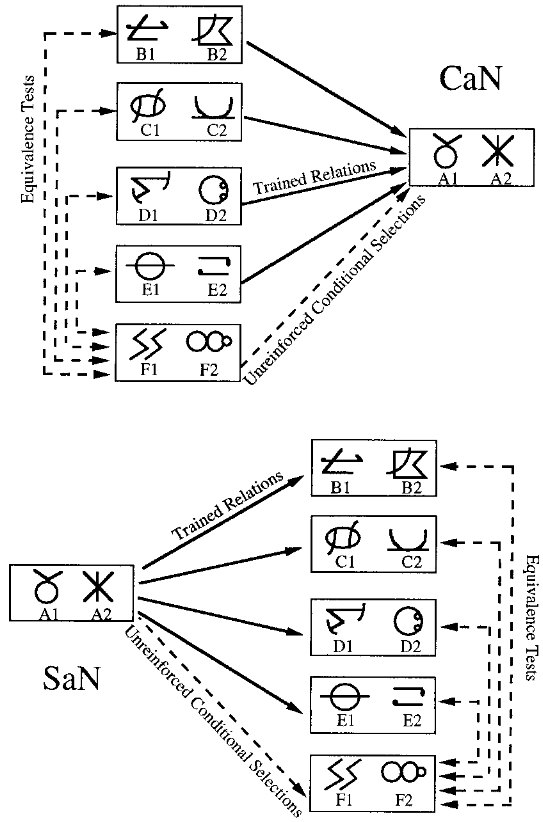
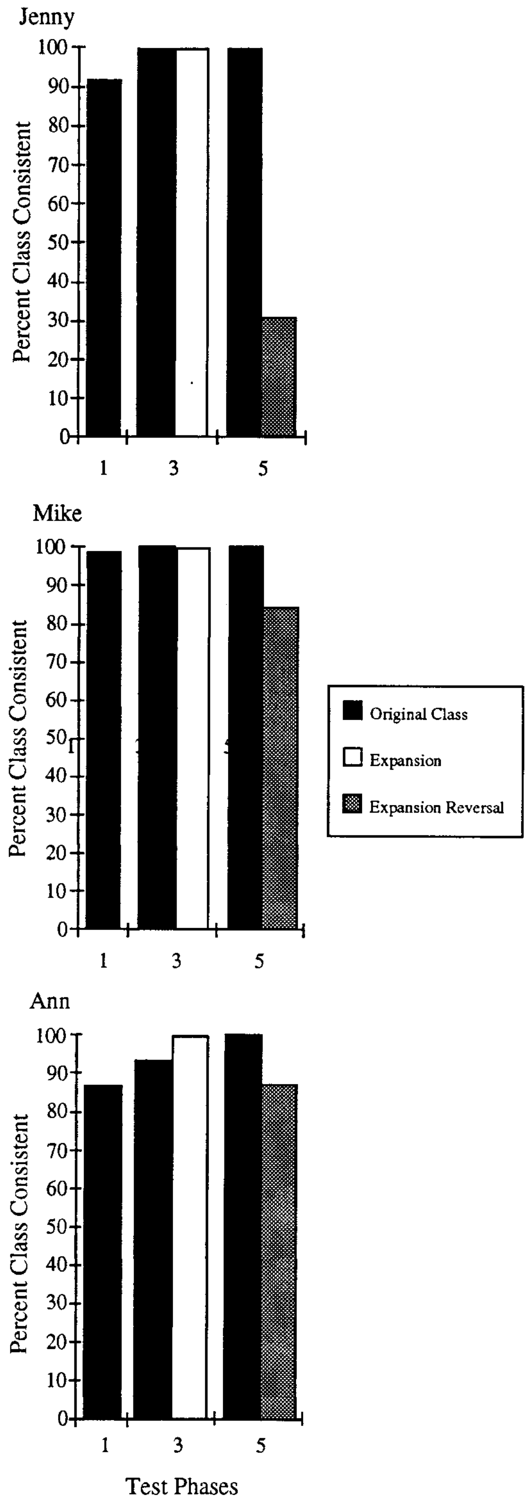


Fig. 5. Schematics of the training structures employed in Experiment 2 to induce class expansion by unreinforced conditional selection. The upper panel depicts the CaN structure leading potentially to two classes expanded to six members each. The lower panel depicts the SaN structure leading potentially to two classes of six members each. Dashed lines indicate the potential relations based on unreinforced conditional selection and the subsequent equivalence tests for class expansion.

Phase 4: Reversal of the Unreinforced Conditional Selection with Reinforcement Procedures

This phase consisted of sessions with trials that were identical to those in Phase 2 for Jenny, Mike, and Ann. An exception for Chet is explained in the Results. In this phase, however, the auditory jingle and token or buzzer followed each response to a comparison stimulus. These consequences were scheduled so that the unreinforced conditional selections made in Phase 2 were now



followed by the buzzer, and the reversed selections (reversal of sample–comparison relations) were followed by the jingle and a token from the experimenter. The same stability criterion applied in Phase 2 was applied in this phase. When the stability criterion was met with scheduled consequences, the consequences were discontinued and the participants were required to meet the stability criterion again in sessions without consequences.

Phase 5: Tests for Equivalence Between the F Stimuli and the Trained Stimuli

In this phase, Jenny, Mike, and Ann were exposed to the same procedures as in Phase 3. An exception for Chet is described in the Results. Four test sessions were conducted in this phase.

RESULTS

The results of Experiment 2 for Jenny, Mike, and Ann are shown in Figure 6. In the final tests of Phase 1 in Experiment 2, class-consistent responding averaged above 90% for Jenny, nearly 100% for Mike, and 87.5% for Ann. For all 3 subjects only two or three sessions of exposure to unreinforced conditional selection procedures were required in Phase 2 to meet the stability criterion. For Jenny and Mike, A1 was consistently selected in the presence of F2, and A2 was selected in the presence of F1. For Ann, F2 was consistently selected in the presence of A1, and F1 was selected in the presence of A2. Class-consistent responding was maintained in four test sessions of Phase 3, demonstrating a maintenance of the original five-member classes and expansion of the original classes to incorporate the F stimuli. Three to six sessions were required in Phase 4 to meet the stability criterion. Following reversal training, there was no disruption of the original five-member

←

Fig. 6. Percentage of class-consistent responding for Jenny, Mike, and Ann from Experiment 2 across three test phases. The black bars represent performance on test trials involving only the original class members. The white bars represent performance on the test trials involving the F stimuli (i.e., class expansion to include the F stimuli) following unreinforced conditional selection. The gray bars represent performance on the test trials involving F stimuli following reversal training.

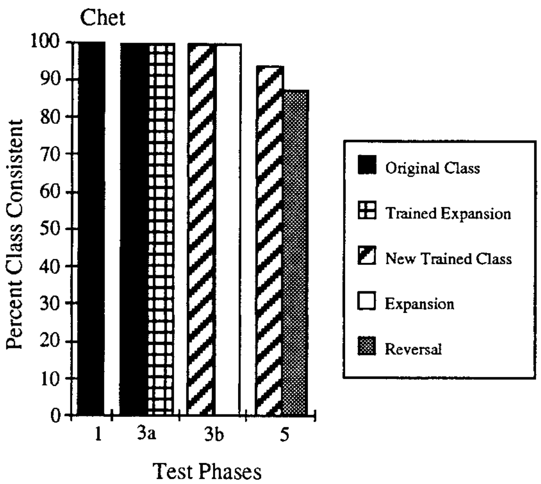


Fig. 7. Percentage of class-consistent responding for Chet from Experiment 2, across four phases. The black bars represent performance on test trials involving only the original class members. For Phase 3a, the second bar in the pair represents performance on test trials involving the trained F relations. For Phases 3b and 5, the diagonally striped bars represent performance on test trials involving the original class expanded to include the F stimuli. The white bar represents performance on test trials involving the G stimuli (i.e., class expansion to include the G stimuli) following unreinforced conditional selection. The gray bars represent performance on the test trials involving G stimuli following reversal training.

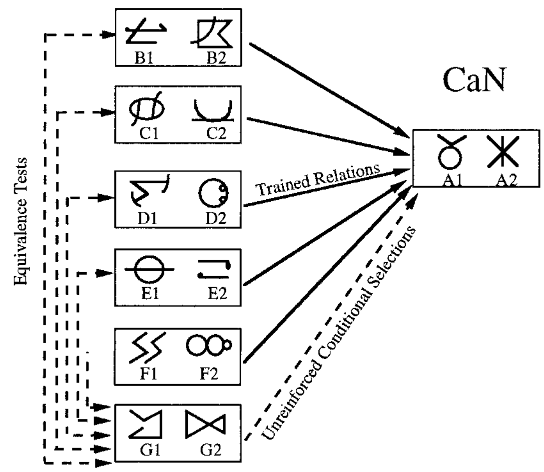


Fig. 8. A schematic of the training structure employed in Experiment 2, Phase 3b, with Chet to induce class expansion by unreinforced conditional selection. Dashed lines indicate the potential relations based on unreinforced conditional selection and the subsequent equivalence tests for class expansion to two seven-member classes.

classes. For Mike and Ann, responses on tests involving F stimuli were generally consistent with class formation based on the reversal training. Jenny's responses on tests involving the F stimuli remained more nearly consistent with the expanded classes based on unreinforced conditional selection than on the expanded classes based on the reversal training. Across the four test sessions of Phase 5, Jenny initially responded in a manner consistent with the reversal, but then in the final three sessions made 11 of 12 responses on tests of the F stimuli that were consistent with the classes based on the previous unreinforced conditional selections.

Figure 7 shows that in the final tests of Phase 1, Chet averaged 100% class-consistent responding. Consistent responding did not occur in Phase 2, however, and by the fifth session a position bias was evident. Thus, Phase 4 training was begun without conducting Phase 3. That is, the decision was made to attempt to expand the classes to six members with training, followed by unreinforced conditional selection procedures with a sixth pair of sample stimuli, rather than proceeding immediately with reversal training. Initially, responses were consistent with Phase 4 training, but when the F stimuli trials were mixed with the original training trials as a preparation for class tests, overall performance deteriorated. After a total of 32 sessions of various forms of refresher training, stable responding was achieved. As shown in Figure 7, Test Phase 3a, the original five-member classes were eventually maintained and expanded to include the F stimuli.

Chet was next exposed to the Phase 2 procedures with a sixth set of sample stimuli. Figure 8 shows the G stimuli that served as samples in the repeat of this phase. Four sessions of exposure to unreinforced conditional selection procedures were required in this phase to meet the stability criteria. A1 was consistently selected in the presence of G2, and A2 was selected in the presence of G1. Class-consistent responding occurred in four test sessions of Phase 3b, demonstrating maintenance of the six-member classes and expansion of those classes to seven stimuli to incorporate the G stimuli. Five sessions were required in Phase 4 to meet the stability criterion. Following reversal training with the G stimuli, responses on tests were usually consistent with class formation based on the reversal training. Overall, 14 of 16 responses on the G tests were consistent with the classes based on reversal, and relations among the other six stimuli in each class were not disrupted.

Discussion

Jenny, Mike, and Ann consistently selected comparison stimuli in the presence of two novel samples or comparisons without trial-by-trial feedback in their first opportunities to do so. Chet did not respond consistently with the first pair of novel stimuli to which he was exposed, but did so with a second novel pair. Thus, some results of this study provide a replication of previous studies by demonstrating a disposition of participants to respond conditionally without trial-by-trial feedback after some exposure to conditional discrimination training. The present study also extends the demonstration of unreinforced conditional selection to younger children and demonstrates for the first time that classes may be expanded as well as merged with this procedure.

With respect to reversal training, the present results with Mike, Chet, and Ann, contrast with those involving larger classes reported by R. Saunders, Saunders, Kirby, and Spradlin (1988), Spradlin et al. (1992), and Pilgrim and Galizio (1995). One difference between the present and previous experiments, however, is that the training structure employed in all prior work produced classes containing three nodal stimuli per class. The structure employed in the present experiment resulted in classes with one node apiece, as did the training reported by Michael and Bernstein (1991) and Pilgrim et al. (1995). Precisely how nodality differences in larger classes could lead to different outcomes is unclear. One possibility is that stimuli in larger classes that become related to more than one node, either by training, unreinforced conditional selection, or by emergent equivalence relations, are more resistant to reorganization than are stimuli related to only one node (in contrast to the prediction of Spradlin et al., 1992, that nodality should not matter). When there is but one node, however, the larger the class the more susceptible it may be to reorganization by the reversal of a single relation (consistent with the prediction of Spradlin et al., 1992).

These structural differences notwithstanding, the results for Jenny in the present experiment were similar to prior results with adults trained with large classes. In the final series of tests, maintenance of the original expanded classes was observed. The relative pliability of responding by the remaining participants with respect to changes in class composition contrasts with the performance disruptions reported previously (Michael & Bernstein, 1991; Pilgrim et al., 1995). The previous studies reporting disruption, however, employed only two conditional discriminations (the minimum number to produce equivalence classes). Reversal of one of these discriminations represents reversal of 50% of the relations that underpin the initial equivalence classes. In the present study, reversal training reversed only 17% (Chet) to 20% (Jenny, Mike, and Ann) of the underpinning relations. Thus, reversal training may not have been disruptive in this instance because so large a proportion of the stimulus–stimulus relations were unaffected by reversal training.

Another explanation for the differences between the results of the present experiment and those of prior research could lie in subject differences. College students and adults with mental retardation may bring repertoires to the experiment that are as yet undeveloped in young children, and these repertoires could override the stimulus controls established by training alone. Verbal repertoires are one such example. Some pattern of covert self-instruction could account for responding in the presence of a particular sample that is consistent with reversal training on one trial and, in the presence of the same sample, is consistent with the original class formation on the next trial. Young children may not be as likely to employ self-instruction. Recent research has demonstrated that preschool children require training to effectively use self-instruction in solving novel conditional discriminations (Grote, Rosales, & Baer, 1996; Grote, Rosales, Morrison, Royer, & Baer, 1997). Experimenter-provided instructions (without stimulus naming) have been implicated in different patterns of results in experiments on stimulus equivalence with college students (e.g., Green, Sigurdardottir, & Saunders, 1991; Sigurdardottir, Green, & Saunders, 1990) and young children (see R. Saunders & Green, 1996). If experimenter-provided instructions can influence results, self-instruction may do so as well.

Another way that participant differences could have produced the differences in results, particularly between those of Pilgrim and Galizio (1995) and the present data, arises from a reconsideration of the data reported by R. Saunders, Saunders, Kirby, and Spradlin (1988) for the 1 participant whose classes were somewhat modified. Maintenance of the original class merger was seen except on test trials that presented the F stimuli, either as the comparisons or as a sample. Responses on the F trials were consistent with reversal training. It is possible that, following an extensive experimental history and a recent reversal history, the presence of the F stimuli functioned as a fifth term in the discriminations (cf. Sidman, 1986). That is, when any trial included an F stimulus in either the sample or comparison position, responding was consistent with the most recent reinforcement history for F-trial stimuli. When any trial did not include F stimuli, responses were consistent with the most recent reinforcement history for these stimuli (when not in the presence of F stimuli). Thus, the results suggest that five-term match to sample might have arisen from reversal training.

A similar analysis can be made with respect to the Pilgrim and Galizio (1995) results (J. M. Saunders, personal communication, August, 1996). Following reversal of their AD and BC relations, it is possible that the presence of both B and C stimuli or both A and D stimuli in a trial set the occasion for responding consistent with the most recent contingencies. Such control should occur on symmetry test trials as well, and Pilgrim and Galizio reported that their participants responded on symmetry tests in a manner consistent with reversal training. On any other trial, such as a BE test for equivalence, only a B stimulus is present; thus, responses might be emitted that are consistent with the most recent reinforcement history for trials in which only one or none of the A, B, C, and D stimuli are present. Pilgrim and Galizio reported that performances on equivalence and transitivity tests were more consistent with the original training. Thus, it can be inferred that the combination of stimuli present on each trial might have acquired control as fifth-term compound (or complex) stimuli, controlling four-term responding on the MTS trials.

Pilgrim and Galizio (1995) concluded with the comment that, ‘‘Collectively, these sorts of findings suggest that equivalence is not a simple by-product of the four-term discriminated operant’’ (p. 237). The present analysis suggests that the four-term analysis of equivalence is not necessarily challenged by any of the reversal data. Rather, four-term contingencies indicative of properties of conditional relations may have come under five-term stimulus control in these experiments. The data from Mike, Ann, and Chet in Experiment 2 show that their responding did not come under five-term control because their test performances became consistent with class organization based on reversal training. Conversely, Jenny's responding was consistent with five-term control, ultimately demonstrating no reorganization of the classes while maintaining reversed performances on the reversed relations.

In summary, those participants who can learn five-term contingencies in the course of the reversal training may do so and show evidence of such learning. Such five-term contingency learning may be closely related to self-instruction, as well as to the age of the participant. Jenny was the oldest participant in the present study (if only by a month), but whether this had any bearing on the differences in her results cannot be determined. Clearly, only further research can resolve the questions and hypotheses that arise from the current data.

GENERAL DISCUSSION

The results of Experiment 1 showed that equivalence classes can be established in young children, even when more than the minimum number of discriminations are taught. In 4 participants younger than 4½ years of age, larger initial classes were established than have previously been demonstrated in participants this young. Overall, the children exhibited learning and performance characteristics similar to adolescents with retardation. That is, equivalence classes were less likely to occur following SaN training than following CaN training.

To date, in research with young children and participants with retardation and with instructions to name stimuli, equivalence classes have been established in 19 of 20 cases following CaN training but in only 3 of 13 cases following SaN training. These results strongly suggest that training structure is a potent factor in whether larger classes develop with these two groups. Yet, in only four of eight cases have larger equivalence classes been established in participants with mental retardation trained within a CaN structure without instructions with stimulus naming. The results suggest that CaN-SaN differences may be greater when instructions are employed than when they are not. Perhaps the apparently greater difficulty in acquiring conditional discriminations linked in a CaN structure prompts naming and naming improves stimulus–stimulus discrimination, leading to better test performance. Perhaps, on the other hand, the lesser difficulty in acquiring the linked SaN discriminations reduces the tendency to name. Either of these variables alone or together may be involved. The database of uninstructed participants is too small at this point for conclusive analysis; more research manipulating these variables will be necessary. Several questions need to be addressed. If the instructions containing names are responsible for the differences, is it the naming of the samples or the naming of the comparisons that is important? Is the answer the same for both SaN and CaN training structures? Is the smaller versus larger class results with SaN structures with young children an instructional effect, a class size effect, or both?

The results of Experiment 2 showed with young children a generalized tendency to respond conditionally in novel conditional discriminations in the absence of scheduled reinforcement contingencies. These results replicate prior results with participants with mental retardation and developmental delay. In the present experiment, the unreinforced conditional selection permitted the expansion of previously developed equivalence classes. Although generally consistent with prior demonstrations of class merger, the present class expansion is the first reported demonstration of this specific effect in any research participant. Training contingencies used to reverse the unreinforced conditional selections were effective in reversing the composition of the previously expanded classes for 3 of 4 participants. Further, the reversal training did not disrupt performance on other discriminations. The present results are important because they demonstrate that some classes can be acquired and subsequently modified without disruption. In many applied situations, such classes clearly have greater practical utility than classes that disintegrate when the membership of a single element is altered or classes that cannot be modified once formed. These results augment the implications of the results of Experiment 1. Instructional strategies involving classification in young children might be enhanced by careful attention to training-structure design. Tentatively, larger classes containing one node would be the optimal training structure if subsequent changes in class membership were anticipated or desired.

Fields and Verhave (1987) listed four parameters that govern the organization of equivalence classes: (a) the number of stimuli per class, (by) the number of nodes, (c) the distribution of nonnodal stimuli among nodal stimuli, and (d) the directionality of training. One aspect of directionality distinguishes CaN training from SaN training: whether the nodal stimulus is the sample stimulus or the comparison stimulus. Fields and Verhave concluded that we can "integrate the results of empirical research by presenting them as functions of the four parameters" (1987, p. 331). Only as more parameters are varied will the full relevance and accuracy of Fields and Verhave's analysis unfold. Subsequent analyses (Barnes, 1994; K. Saunders et al., 1993; and see Footnote 1) appear to be explaining *why* Fields and Verhave's parameters are important and how they affect learning, leading to different results. Further, the influences of experimenter-provided instructions and stimulus names on Fields and Verhave's parameters continue to be variables that are ripe for empirical examination and theoretical discussion (see the *Journal of the Experimental Analysis of Behavior, 65,* 183–353).

REFERENCES

Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of training design. *The Psychological Record, 47,* 309–320.

Barnes, D. (1994). Stimulus equivalence and relational frame theory. *The Psychological Record, 44,* 91–124.

Barnes, D., Browne, M., Smeets, P., & Roche, B. (1995). A transfer of functions and a conditional transfer of functions through equivalence relations in three- to six-year-old children. *The Psychological Record, 45,* 405–430.

Barnes, D., McCullagh, P. D., & Keenan, M. (1990). Equivalence class formation in non-hearing impaired children and hearing impaired children. *The Analysis of Verbal Behavior, 8,* 19–30.

Brady, N. C., & Saunders, K. J. (1991). Considerations in the effective teaching of object-to-symbol matching. *Augmentative and Alternative Communication, 7,* 112–116.

Carrigan, P. F., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 58,* 183–204.

Carter, E. E., & Eckerman, D. A. (1975). Symbolic matching by pigeons: Rate of learning complex discriminations predicted from simple discriminations. *Science, 187,* 662–664.

Carter, E. E., & Werner, T. J. (1978). Complex learning and information processing by pigeons: A critical analysis. *Journal of the Experimental Analysis of Behavior, 29,* 565–601.

Devany, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled participants. *Journal of the Experimental Analysis of Behavior, 46,* 243–257.

Dixon, M. H., & Spradlin, J. E. (1976). Establishing stimulus equivalence among retarded adolescents. *Journal of Experimental Child Psychology, 21,* 144–164.

Dixon, L. S. (1977). The nature of control by spoken words over visual stimulus selection. *Journal of the Experimental Analysis of Behavior, 27,* 433–442.

Dugdale, N., & Lowe, C. F. (1990). Naming and stimulus equivalence. In D. E. Blackman & H. Lejeune (Eds.), *Behaviour analysis in theory and practice: Contributions and controversies* (pp. 115–138). Hillsdale, NJ: Erlbaum.

Eikeseth, S., & Smith, T. (1992). The development of functional and equivalence classes in high-functioning autistic children: The role of naming. *Journal of the Experimental Analysis of Behavior, 58,* 123–133.

Fields, L., Hobbie, S. A., Adams, B. J., & Reeve, K. F. (in press). Effects of training directionality and class size on equivalence class formation by adults. *The Psychological Record.*

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior, 48,* 317–332.

Fields, L., Verhave, T., & Fath, S. (1984). Stimulus equivalence and transitive associations: A methodological analysis. *Journal of the Experimental Analysis of Behavior, 42,* 143–158.

Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 229–262). New York: Plenum.

Green, G., Sigurdardottir, Z. G., & Saunders, R. R. (1991). The role of instructions in the transfer of ordinal functions through equivalence classes. *Journal of the Experimental Analysis of Behavior, 55,* 287–304.

Grote, I., Rosales, J., & Baer, D. M. (1996). A task-analysis for the shift from teacher instructions to self-instructions in performing an in-common task. *Journal of Experimental Child Psychology, 63,* 339–357.

Grote, I., Rosales, J., Morrison, K., Royer, C., & Baer, D. M. (1997). A use of self-instruction to extend the generalization of a self-instructed in-common discrimination. *Journal of Experimental Child Psychology, 66,* 144–162.

Johnson, C., & Sidman, M. (1993). Conditional discrimination and equivalence relations: Control by negative stimuli. *Journal of the Experimental Analysis of Behavior, 59,* 333–347.

Lazar, R. M., Davis-Lang, D., & Sanchez, L. (1984). The formation of visual stimulus equivalences in participants. *Journal of the Experimental Analysis of Behavior, 41,* 251–266.

Lowe, C. F., & Beasty, A. (1987). Language and the emergence of equivalence relations: A developmental study. *Bulletin of the British Psychological Society, 40,* A42.

Mackay, H. A. (1991). Stimulus equivalence: Implications for the development of adaptive behavior. In R. Remington (Ed.), *The challenge of severe mental handicap: An applied behaviour analytic approach* (pp. 235–259). New York: Wiley.

McIlvane, W. J., & Dube, W. V. (1996). Naming as a facilitator of discrimination. *Journal of the Experimental Analysis of Behavior, 65,* 267–272.

Michael, R. L., & Bernstein, D. J. (1991). Transient effects of acquisition history on generalization in a matching-to-sample task. *Journal of the Experimental Analysis of Behavior, 56,* 155–166.

Pilgrim, C., Chambers, L., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: II. Children. *Journal of the Experimental Analysis of Behavior, 63,* 239–254.

Pilgrim, C., & Galizio, M. (1995). Reversal of baseline relations and stimulus equivalence: I. Adults. *Journal of the Experimental Analysis of Behavior, 63,* 225–238.

Saunders, K. J. (1989). Naming in conditional discrimination and stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 51,* 379–384.

Saunders, K. J., Saunders, R. R., Williams, D. C., & Spradlin, J. E. (1993). An interaction of instructions and training design on stimulus class formation: Extending the analysis of equivalence. *The Psychological Record, 43,* 725–744.

Saunders, K. J., & Spradlin, J. E. (1990). Conditional discrimination in mentally retarded subjects: The development of generalized skills. *Journal of the Experimental Analysis of Behavior, 54,* 239–250.

Saunders, R. R., & Green, G. (1992). The nonequivalence of behavioral and mathematical equivalence. *Journal of the Experimental Analysis of Behavior, 57,* 227–241.

Saunders, R. R., & Green, G. (1996). Naming is not (necessary for) stimulus equivalence. *Journal of the Experimental Analysis of Behavior, 65,* 312–314.

Saunders, R. R., Saunders, K. J., Kirby, K. C., & Spradlin, J. E. (1988). The merger and development of equivalence classes by unreinforced conditional selection of comparison stimuli. *Journal of the Experimental Analysis of Behavior, 50,* 145–162.

Saunders, R. R., Wachter, J. A., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior, 49,* 95–115.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research, 14,* 5–13.

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213–245). Hillsdale, NJ: Erlbaum.

Sidman, M. (1987). Two choices are not enough. *Behavior Analysis, 22,* 11–18.

Sidman, M. (1990). Equivalence relations: Where do they come from? In D. E. Blackman & H. Lejeune (Eds.), *Behavior analysis in theory and practice: Contributions and controversies* (pp. 93–114). Hillsdale, NJ: Erlbaum.

Sidman, M. (1994). *Equivalence relations and behavior: A research story.* Boston: Authors Cooperative.

Sidman, M., Kirk, B., & Willson-Morris, M. (1985). Six-member stimulus classes generated by conditional-discrimination procedures. *Journal of the Experimental Analysis of Behavior, 43,* 21–42.

Sidman, M., & Tailby, W. (1982). Conditional discriminations vs. matching-to-sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37,* 5–22.

Sidman, M., Willson-Morris, M., & Kirk, B. (1986). Matching-to-sample procedures and the development of equivalence relations: The role of naming. *Analysis and Intervention of Developmental Disabilities, 6,* 1–20.

Sidman, M., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior, 52,* 261–274.

Sigurdardottir, Z. G., Green, G., & Saunders, R. R. (1990). Equivalence classes generated by sequence training. *Journal of the Experimental Analysis of Behavior, 53,* 47–63.

Spradlin, J. E., Cotter, V. W., & Baxley, N. (1973). Establishing a conditional discrimination without direct training: A study of transfer with retarded adolescents. *American Journal of Mental Deficiency, 77,* 556–566.

Spradlin, J. E., Saunders, K. J., & Saunders, R. R. (1992). The stability of equivalence classes. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 29–42). Reno, NV: Context Press.

Spradlin, J. E., & Saunders, R. R. (1986). The development of stimulus classes using match-to-sample procedures: Sample classification vs. comparison classification. *Analysis and Intervention in Developmental Disabilities, 6,* 41–58.

Stromer, R., & Osborne, J. G. (1982). Control of adolescents' arbitrary matching-to-sample by positive and negative stimulus relations. *Journal of the Experimental Analysis of Behavior, 37,* 329–348.

Urcuioli, P. J., & Zentall, T. R. (1993). A test of comparison-stimulus substitutability following one-to-many matching by pigeons. *The Psychological Record, 43,* 745–759.

Urcuioli, P. J., Zentall, T. R., Jackson-Smith, P., & Steirn, J. N. (1989). Evidence for common coding in many-to-one matching: Retention, intertrial interference, and transfer. *Journal of Experimental Psychology: Animal Behavior Processes, 15,* 264–273.

Wetherby, B., Karlan, G. R., & Spradlin, J. E. (1983). The development of derived stimulus relations through training in arbitrary-matching sequences. *Journal of the Experimental Analysis of Behavior, 40,* 69–78.

Received November 12, 1997
Final acceptance November 10, 1998